Session 3

CUSTOMER SURVEYS

Quality Management for Customer Satisfaction Surveys

Richard M. Devens, Jr.
Executive Editor
Monthly Labor Review
U.S. Bureau of Labor Statistics

Introduction.

This is the story of a customer satisfaction survey done for the Employment and Unemployment Statistics Quality Council at the Bureau of Labor Statistics. This is was the program's first customer satisfaction survey, and we are still learning from it. What I hope to pass on in this paper are the lessons learned about serving the "other customers," the executives that sponsored the survey and the front-line staff at the survey's focus. In other words, how "fit for use" was the National Survey of Users of Employment and Unemployment Statistics? Before plunging into that, I will take a few minutes to explore the importance of customer surveys and to outline the technical process of designing and conducting this one.

The reason for conducting any customer satisfaction survey is the position customers hold in the guiding principles of total quality management (TQM):

- customer focus,
- employee involvement,
- continuous improvement.

Customer focus, in my mind, is both the most important of these principles and the most difficult to persuade many public-sector managers to accept.

Continuous improvement is normally accepted straight off, usually with the assertion that the organization is already practicing it. The TQM purist might quibble that managers often mean their organization is always on the lookout for the big breakthrough, rather than practicing Deming's Fifth Point. [Improve constantly and forever every process for planning, production, and service.] My own observation has been that managers really do want to improve their operations, one way or another.

Employee involvement is a bit harder to sell. Many executives are used to and, quite frankly, happy with a command-and-control structure. In the case of the statistical agencies, such organizations were tremendously successful at organizing the armies of data collectors, mail room clerks, document controllers, coders, key entry workers, data reviewers, statistical assistants, statistical typists, junior economists, computer operators, computer programmers, research assistants, supervisory statisticians, senior economists, printers, and Assistant Commissioners that it took, and still takes to a fair extent, to produce a few tables of accurate, timely, relevant numbers.

Whether these hierarchies will work as well when data collection becomes automated, databases are connected through electronic data interchange (EDI), and performance becomes more dependent on the commitment of highly-skilled, self-confident, and very independent professionals is the issue. I believe such developments will lead organizations to embrace employee involvement models sooner rather than later--and most executives realize it, however grudgingly.

Customer focus, in contrast, is a very difficult concept for public-sector managers to accept at all, let alone embrace. The first reaction is, "We don't sell anything, so we don't have customers." Even after getting over this "filthy lucre" barrier, there is, especially in "craft" or "engineering" cultures such as those of the statistical bureaus, a deep skepticism about the fitness of the customer to make rational decisions or even to know what they want. These reactions are evident deep down into the structure of such agencies. Where the first-line will quickly accept the notion of getting involved in and taking greater responsibility for technical improvement, there is little enthusiasm for treating their work as a customer-satisfying process, not an estimates- or analysis-producing process.

The upshot of all this for the manager of a customer satisfaction survey is that there are two other--and perhaps more difficult --customers that must be considered in parallel with the external customer: the executive-level sponsors and the front-line staff. The rest of this paper overviews the National Survey of Users of Employment and Unemployment Statistics and its findings, the interaction of the project with its sponsors, the interaction of the project with the front-line staff, and the reactions of these "other" customers to the survey.

Outline of the National Survey of Users of Employment and Unemployment Statistics

The National Survey of Users of Employment and Unemployment Statistics is based on the premise that customer satisfaction is measured by the discrepancy between the client's needs and expectations and the client's perception of our performance. In the marketing literature, this is known as "disconfirmation" theory. The survey measures expectations and performance in five broad factors:

- Data quality: The accuracy, relevance, and timeliness of our statistics.
- Tangibles: The appearance and understandability of our materials.
- Dependability: Our demonstrated ability to perform promised services reliably, correctly, and promptly.
- Assurance: The knowledge of our employees and their ability to convey trust and confidence.
- Empathy: The caring, courteous, individualized attention we provide.

Each factor is represented by specific statements in the questionnaire. (See box.) The questionnaire also provides for an independent ranking of the importance of the factors and for general evaluations of satisfaction with our statistics and associated services.

Quality Factors and Their Proxies
(Question number in parentheses)

Data Quality

(2) The demographic, geographic, and industrial coverage of the statistics is sufficient for my needs.

(7) The data provided meet my standards of accuracy and reliability.

(8) The data provided meet my standards of timeliness and currency.

Dependability

(1) Staff are always available during their normal working hours.

(4) My questions are answered promptly and dependably.

(5) It is easy to get in touch with someone who can answer my questions.

(14) The information I ask for is sent in the medium and format requested.

Tangibles

(6) Materials provided make sense and can be understood without additional information.

(14) The information I ask for is sent in the medium and format requested.

Assurance

(9) Staff are knowledgeable and competent.

(11) Staff can clearly explain conceptual and analytical issues without using overly technical language.

(12) Staff can clearly explain the technical limitations of the data.

Empathy

(3) Staff make me feel that I can call back for additional clarification or data.

(4) My questions are answered promptly and dependably.

(10) Staff are courteous.

(13) Staff go out of the way to understand and fulfill my requests.

(15) Apologies are rendered for inconveniences such as delay or misunderstanding of my needs.

Clients rated the statements on the quality of our statistics and services on 5-point scales for their expectations of quality and their perception of our performance. The expectation score is subtracted from the performance score to yield the "performance gap" for any specific statement. The performance gap for a factor is the mean gap for the set of statements that represent it.

In addition to the customer satisfaction scales, the survey asks how clients use our data, which programs they have utilized, and what channels of distribution were used to access data. We also provided space for comments.

Designing the Survey

We developed this user-friendly questionnaire using cognitive research methods including focus groups, think-aloud interviews, and a pilot test. Each of these methods identified errors and we were able to take corrective action before taking the final survey into the field.

In the field, Dillman's Total Design Method was followed closely, with the exception of experimental variations in the third and final follow-ups. Clients selected for the survey received several mailings:

- A notice arrived at the customer's address a few days before the primary questionnaire package.
- A thank-you/reminder letter followed the questionnaire by about a week.
- A second package went out two weeks after the "tickler."
- Final prompting, experimentally split between certified mail and telephone prompts, began 2 weeks after that.

This intensive data collection methodology yielded a usable response rate of 87.8 percent.

Two minor modifications to the Dillman method were necessary. First, the front cover of the questionnaire was not illustrated with graphics because of the limited space, and the stationery size was the ordinary 8 1/2 by 11. Second, the reminder/thank-you postcard was replaced by a reminder/thank-you letter because in-house constraints allowed letter production only.

The experimental exercise conducted in the third follow-up tested certain refinements to the Total Design Method for use in the establishment setting. Two weeks after the second follow-up, each of the remaining nonrespondents was randomly assigned to either certified mail follow-up or telephone prompting. In the control group, "holdouts" received the third follow-up packet by certified mail containing a replacement questionnaire, a business reply envelope, and a cover letter. The wording of the cover letter was different from the cover letters used in the preceding follow-ups; we softened and relaxed the wording but emphasized explanations of why this additional follow-up is important and is sent by certified-mail.

In the treatment group, nonrespondents were contacted by trained, experienced telephone prompters. We prepared a survey-specific training agenda, drawing on insights from nonresponse

conversion efforts in telephone follow-up surveys. The training included practice of scripted telephone procedures including appropriate reactions to specific reasons for refusal, discussion of persuasive techniques, and use of call record sheets. Approaches to locate the sample subject and find the best time to call back were also included in the training.

The survey's sample frame was constructed from two sources. First, client contact staff in the National program offices, the Regional Offices, and the Inquiries and Correspondence section logged contacts during September-November 1992. Program managers and senior executives provided separate lists of "regular" clients--persons maintaining on-going professional contact with our programs.

The lists were merged and duplicate entries removed. The resulting sample frame contained 3553 names which were stratified based on the program office that was the point of contact. Two additional strata were formed: one for all of the regular clients and another for all of the customers who were logged in by more than one program. The total sample of 999 clients was obtained by selecting samples of approximately equal size from all strata except one. Members of the stratum of regular users were included in the sample with certainty.

The response rate figures from the national survey of users of Employment and Unemployment Statistics are shown in Table 1. After the second replacement questionnaire mailout, the overall response rate had already reached 75%, which is the average overall response rate for TDM-based surveys. The third and final follow-up boosted the response rate by 13 percentage points to approximately 88%.

Table 1. Response and Conversion Rates

Mailing	Conversion Rate (%)	Overall Rate (%)	N*
Prenotice (Day 1)			999
1st mailout (Day 8)	28.65	28.65	998
Reminder (Day 15)	50.87	65.24	978
2nd mailout (Day 29)	28.53	75.77	970
3rd follow-up (Day 43)			
Certified mail	48.21		
Phone Prompting	44.44		
Overall	46.45	87.68	950
Close-out (Day 71)			

* The sample size declined as ineligibles were uncovered through the data collection process.

Summary of Findings. Despite averaging 4.08 out of a possible 5 points on the performance scale of our survey, we did not fully meet our customers' expectations. (Expectations averaged 4.46 out of 5.)

Considering the major factors displayed on chart 1, our "performance gaps"--the average difference between our performance and our customers' expectations across the statements that represented the factors--were:

1.	Data quality	(-0.66)
2.	Dependability	(-0.52)
3.	Tangibles	(-0.34)
4.	Assurance	(-0.28)
5.	Empathy	(-0.17)

Using an expectations/performance grid --a "customer window" in the most recent jargon-- to analyze individual statements shows specific areas to concentrate our efforts on. (See chart 2.) In this graphic display, the intersection of the axes represents the grand means for customers' expectations (Y-axis) and their perception of our performance (X-axis). The points plotted for each statement are the ordered pair of Z-scores. According to this analytical tool, the important places to "Concentrate" corrective strategies are:

1. More timely (#8) and detailed (#2) data
2. Making it easier to find someone to answer your questions (#5)
3. Providing clearer materials (#6)

The survey report expressed these in terms of three strategic themes for improvement:

- Get Faster: Make statistical products available to the public more quickly.
- Basic Service First Time/Full Service Every Time: Have analysts able to answer broader ranges of inquiries, rather than transfer customers across program lines.
- Clarity, Clarity, Clarity. Make products and services easier to understand from the customer's point of view.

Chart 1

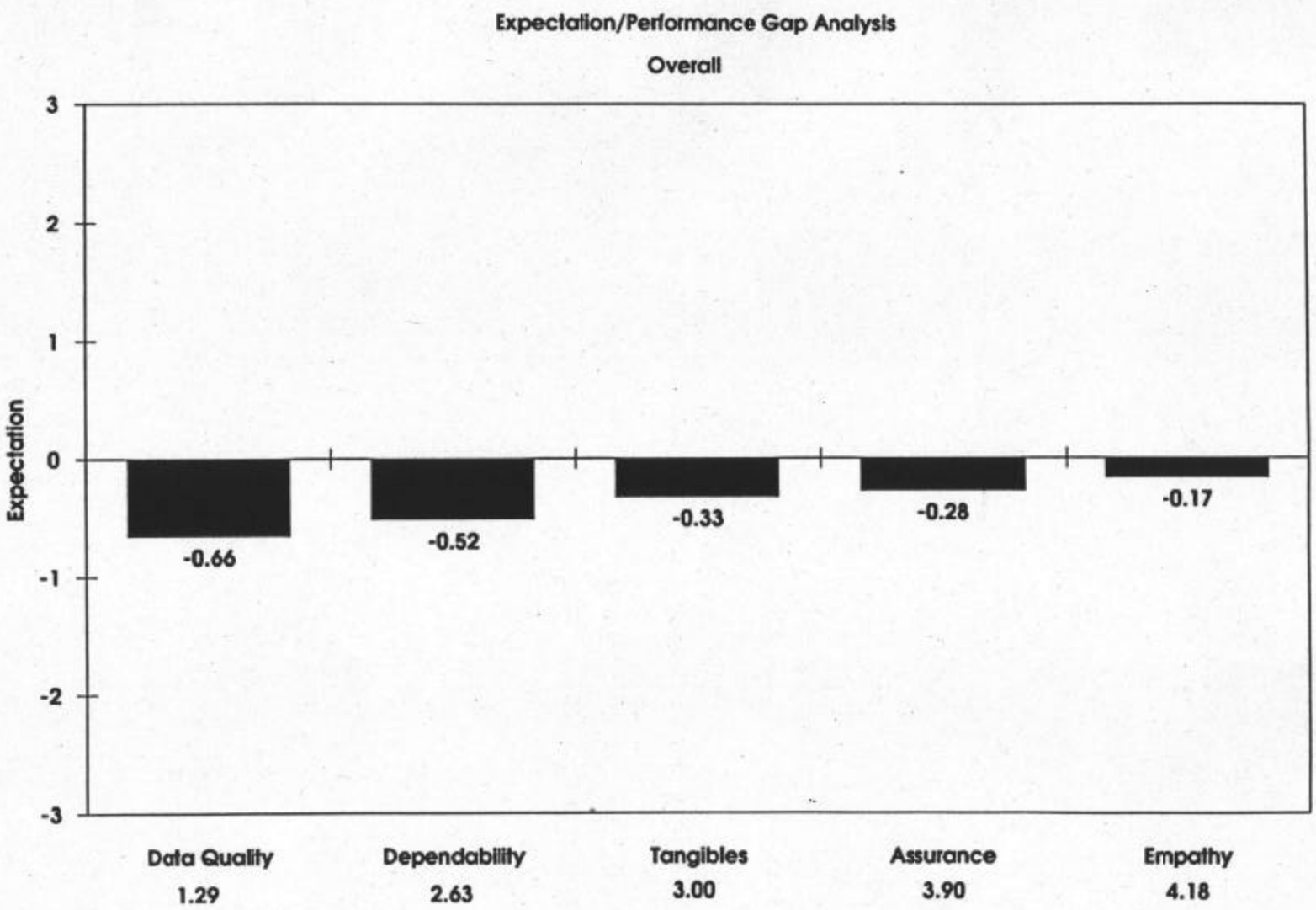
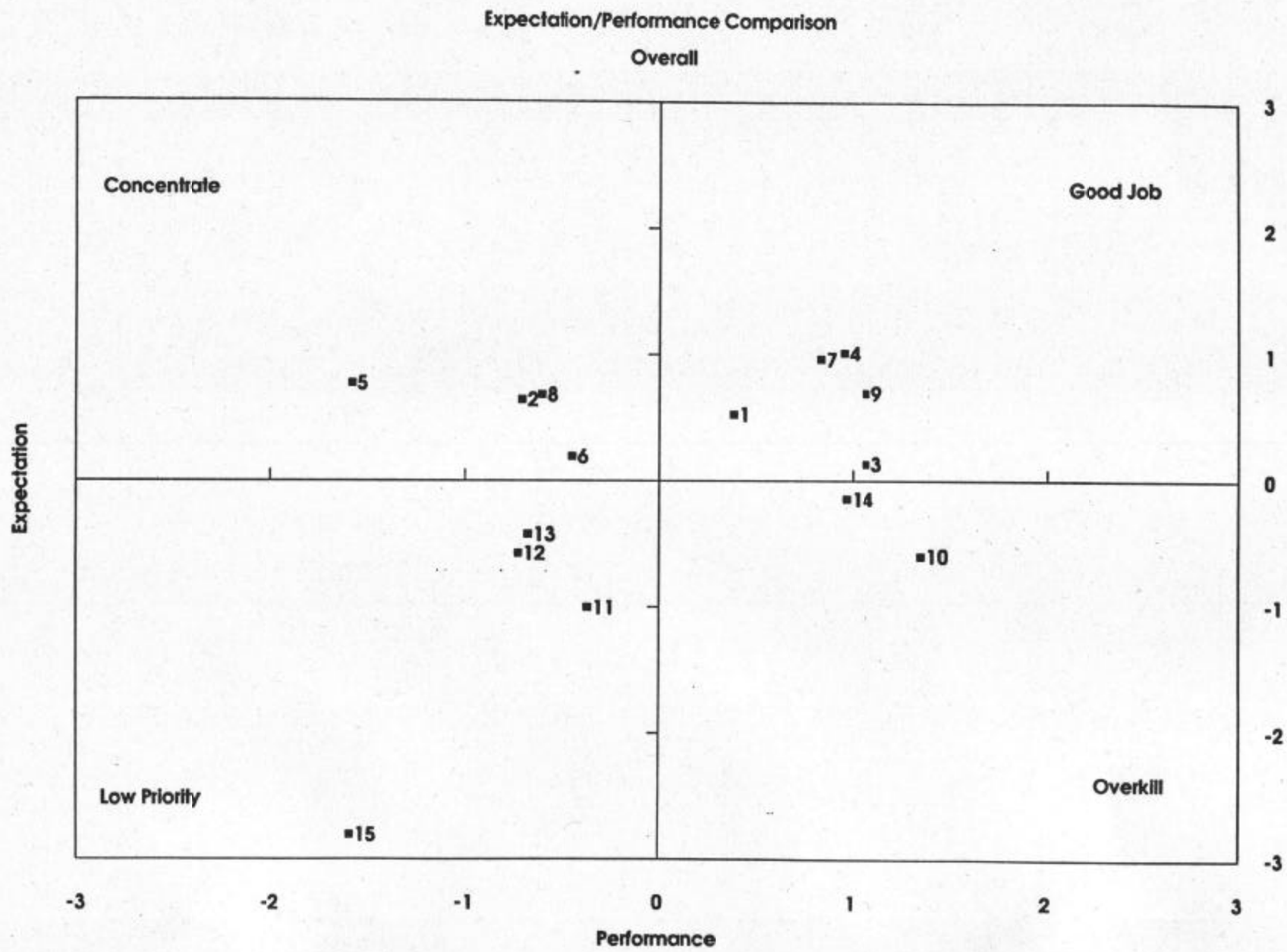

Expectation/Performance Gap Analysis
Overall
Expectation
3
2
1
0
-1
-2
-3
-0.66
-0.52
-0.33
-0.28
-0.17
Data Quality
1.29
Dependability
2.63
Tangibles
3.00
Assurance
3.90
Empathy
4.18

Chart 2


Expectation/Performance Comparison
Overall
Concentrate
Good Job
Low Priority
Overkill
Expectation
Performance
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
5
2
8
7
4
9
1
6
3
14
13
12
10
11
15

Interaction with executive-level sponsors. The National Survey of Users of Employment and Unemployment Statistics is an "infrastructure" project on the part of the Employment and Unemployment Statistics (EUS) Quality Council. The Council itself is the executive-level coordinating body for the EUS Quality Improvement Program. The Council took its first look at conducting a customer survey at its August 1991 meeting. I was assigned the task of pulling together a few ideas on how such surveys were conducted and how they might apply to the EUS quality program. At their November 1991 meeting, the Council approved moving forward.

For the next several months, the minutes don't do the meetings justice. "..council members expressed concern," "..requested that a full proposal be prepared before proceeding," "...discussed the difficulties...." "After some discussion, it was agreed...." I am sure all of you are aware of what lurks behind these bland formulations.

After laboriously negotiating final approval, the survey team administered the instrument to the Quality Council itself, with the instruction, "Complete the questionnaire as if you were the 'average' customer." This exercise had the twin goals of further educating the Council on the survey and developing a baseline measure of the Council's awareness of customer needs and knowledge of the customers' perceptions of our performance.

As a baseline exercise, the Quality Council identified the same order of priority among the major quality factors that customers did. Interestingly enough, however, the absolute sizes of the perceived performance gaps were actually larger among Quality Council members than among customers. The primary source of the larger gaps, as shown below, was lower performance ratings by the Quality Council.

	Customers	**Council**
Average performance	4.08	3.80
Average expectation	4.46	4.36

How this exercise worked as an educational tool is a good question. My subjective evaluation is that the Council members themselves perceived it fairly narrowly in its baseline setting role, and would be surprised to find out about its covert objective of preparing them to more fully understand the survey's results.

When the final report of the project was drafted, the Quality Council was briefed on its contents and provided with copies for comment and approval. Once approval was obtained, final reports with more extensive technical documentation were published and circulated to the Council and the staff of the employment and unemployment statistics activity. The Business Research Advisory Council to the Bureau also expressed an interest in the survey and its results. Members of the Quality Council attended that briefing as well.

Interaction with customer-contact staff. Another set of customers for the survey and its results is the front line customer-contact staff. This is the group of our colleagues that provided roughly 3,500 customer names and addresses representing over 5,000 direct contacts over a 3-month period. These are also the people whose work product was put under scrutiny by the survey and the upon whom much of the burden of improvement would be likely to fall.

From the outset of the survey, five senior professionals from customer-contact units were assigned to the project. Their substantive contributions were critical and they also served as a "backchannel" of informal feedback between the staff and the survey team. That channel was, during the universe-building phase, our best means of helping the staff focus on keeping a complete log of contacts. (As a result of our debriefing of the representatives, one improvement we are likely to make in future surveys is a shorter log-in.) While the survey was in the field, the backchannel kept the staff informed about our progress.

Other interactions with the customer-contact staff included formal training sessions on the objectives, concepts, and methods of the survey and the procedures they would follow maintaining the universe log. In addition, the log procedures were documented on the forms themselves along with explanations of the purposes of the survey itself and of some of the most critical pieces of universe information--e.g., telephone numbers.

In general, interaction with this group of customers is something we should improve on. Some regional office information staffs had virtually no training or documentation of the survey or their role in it until what might be generously called the last minute. Interaction with the Inquiries and Correspondence Branch of the Office of Publications, while more timely, never reached the extent or intensity needed.

Reaction of sponsors. The most important measure of the success of a project such as the National Survey of Users of Employment and Unemployment Statistics is the action it prompts the organization to take. On this score the results are promising, but not overwhelming. Actions "moved onto higher priority time paths" to improve on the critical data quality factor include:

- Advancing the review and release of State-wide Local Area Unemployment Statistics (LAUS) data by 2-3 weeks (from a baseline of roughly 9 weeks after the reference period).
- Converting 20,000 late respondents to the Current Employment Statistics (CES) survey to automated self-reporting using an advanced touch-tone telephone data collection technology to improve the timeliness of these reports.
- Hosting the International Occupational Classification Conference to provide a forum for discussing new ideas and alternative approaches to the details of occupational categorization.
- Expanding service-sector detail for Current Employment Statistics by adding 108 new series to our most detailed publication and 20 seasonally-adjusted series to the employment news release.

In general, these have been projects that were on various burners to start with--the most the survey can claim is that some were completed more quickly and with more fanfare. There have been a few initiatives to start addressing some of the issues of service quality:

- Increasing the clarity of hard copy information sent to customers
- Resource book for information calls
- Developing new-employee training module for customer service
- Cross-program briefing on data availability.

My personal evaluation of the impact of the survey is that it was useful, but not nearly in proportion to the skills exercised or to the resources expended. That sense of disproportionality of effort leads me to the point of the paper--how well served are the "other" customers?

User-survey-users' surveys. To find out, I conducted a pair of informal surveys of the two groups of "other" customers. The survey of the Quality Council asked for their evaluation of the importance of the strategic directions the results pointed to, an evaluation of the communication processes between the Council and the project team, and an evaluation of the team's effectiveness at communicating the results. A similar survey was conducted among the front-line staff.

The results of the executive survey indicated that the group found that they rated the importance of the 3 strategic themes quite closely together between 5.3 and 5.8 on a scale of 1-to-7 (Not important at all to Extremely important). The highest score went to the Clarity, Clarity, Clarity theme.

The executives' evaluations of the effectiveness of our communication of the concepts and methods of the survey and of the results of the survey were devastatingly frank. On 5-point scales, the scores were 3.43 on effective communication of concepts (between "fairly" and "very" effective) and 3.57 on clear communication of results (again between "fairly" and "very"). These low scores, and remember they came from colleagues, probably reflect the reason the response to the survey was not overwhelming--the credibility of the product was not established and the results were not clearly communicated to the sponsors.

Reaction of front-line staff. The front-line's reaction to the survey is neatly summarized by the response rate to the survey included in the individual copies of the final report--almost 8 percent. Obviously, our efforts to engage this group fell short. The open question is why did it happen? I fear that the real reason was a fundamental failure to convince the front line that the customer satisfaction survey was serious. This may be the most significant quality issue for the National Survey of Users of Employment and Unemployment Statistics.

For what it is worth, the front-line survey found that among six respondents the importance ratings of the themes ranged from 4.7 to 5.7, with the clarity issue highest once again. The scores for effective, clear presentation of the concepts, methods and results of the customer survey--3.2 for effective presentation of objectives, concepts, and methods and 3.3 on clear communication of the results-- were even lower than those given by the sponsors.

Conclusions. Customer satisfaction is the "outcome" of any statistical or information service. This must often be measured quite separately from the "output" of programs. Output measures too often tell more about what is important to us than what is important to the customer. We in the Federal statistical community have always been concerned about hard measures of the output, "accurate data". We have only just now become aware that the soft outcome, "satisfaction with promptness and dependability of service," is perhaps even more important. That is why customer satisfaction surveys are useful--they are tools to measure and manage that outcome.

To be taken seriously as management tools, however, customer satisfaction surveys must be credible to the "other customers"--the people who should respond to the results. My point is that to obtain that outcome, the customer survey manager must establish credibility in advance and not think that good output--a clever report based on sound data--will suffice.

Note: All material in this article is solely the reponsibility of the author. The views expressed here do not necessarily reflect the policy of the Bureau of Labor Statistics or the views of other BLS staff members.

COMPARABILITY IN CUSTOMER SATISFACTION SURVEYS:

PRODUCTS, SERVICES, AND GOVERNMENT AGENCIES

by

Michael D. Johnson

National Quality Research Center
The University of Michigan
School of Business Administration
Ann Arbor, MI 48109-1234

Phone: 313-764-1259
Fax: 313-763-9768

April, 1994

To appear in *New Directions in Statistical Methodology*, Washington, DC: Office of Management and Budget, forthcoming. Do not quote or reproduce without permission. The author thanks the Swedish Post Office, the National Quality Research Center at the University of Michigan's School of Business Administration, and the International Center for Studies of Quality and Productivity at the Stockholm School of Economics for providing the data described in the paper.

COMPARABILITY IN CUSTOMER SATISFACTION SURVEYS: PRODUCTS, SERVICES, AND GOVERNMENT AGENCIES

ABSTRACT

This paper describes recent advances in customer satisfaction surveys and their implications for government agencies. Many agencies are in the process of implementing customer satisfaction monitoring systems and identifying appropriate private sector benchmarks. Satisfaction models and survey methods currently being used to produce national customer satisfaction indices are described. These efforts illustrate a number of important steps that should help government agencies produce meaningful measures of satisfaction and identify private sector industries that provide realistic agency benchmarks.

INTRODUCTION

Customer satisfaction has emerged as an important benchmark for gauging the performance of various economic agents over the past decade. Manufacturers of durable and nondurable products, retailers, service providers, utilities, and government agencies alike have implemented, or are in the process of implementing, customer satisfaction measurement systems. At a micro-level, these systems monitor a firm's or agency's primary asset - their customers - and provide important diagnostic information needed to improve or maintain satisfaction. At a macro-level, Sweden and Germany have implemented national customer satisfaction indices to monitor the major sectors of their economies while the United States, Taiwan, and New Zealand are in the process of doing the same.

There is disagreement, however, among psychologists, economists, consumer researchers, public policy makers, and others regarding the merits of comparing satisfaction across individual and industries. The ever broadening arena of customer satisfaction, in conjunction with recent advances in how satisfaction is surveyed and operationalized, shed light on this long-standing debate. The goal of this paper is to describe these developments and discuss their implications for government agencies. Many agencies now find themselves, for the first time, asking such questions as, "who are the customers served by our agency, who should be our customers, what standards should we use, and what comparable businesses should we benchmark on?"[1] Recent advances in satisfaction survey methods provide important insights into how government agencies should survey their customers. Sweden's experience with a national Customer

---

[1] Section 1 of President William J. Clinton's executive order, dated September 11, 1993, begins as follows:

> In order to carry out the principles of the National Performance Review, the Federal Government must be customer-driven. The standard of quality for services provided to the public shall be: Customer service equal to the best in business. For the purposes of this order, "customer" shall mean an individual or entity who is directly served by a department or agency. "Best in business" shall mean the highest quality of service delivered to customers by private organizations providing a comparable or analogous service.

Satisfaction Barometer (the SCSB) also illustrates which private sectors businesses provide benchmarks or standards of comparison for these agencies. First, however, the nature of the debate over the comparability of satisfaction and the issues involved are described.

SATISFACTION AND THE HAPPY SLAVE PROBLEM

Customer satisfaction is a customer's evaluation of their overall experience with a product or service to date (Johnson and Fornell 1991; Johnson et al. 1994). This definition of satisfaction is consistent with existing views in economic psychology, where satisfaction is often equated with notions of subjective well-being (Van Raaij 1981), and economics, where satisfaction is equated with post-purchase consumption utility (Meeks 1984). Because it describes the customer's total consumption experience, satisfaction predicts customer loyalty and a firm's subsequent "profitability." In the private sector this "profit" is bottom line return on assets (Anderson et al. 1994). For government agencies, the benefits of increased customer satisfaction range from budget considerations, to more efficient use of taxpayer dollars, to the creation of a more positive image, to compliance (e.g., for the Internal Revenue Service).

There is a long standing debate in economics over the comparability of satisfaction across individuals and industries (see Hammond 1991 for a review and extensive bibliography). Bentham (1802) defended the comparability of satisfaction as both possible and necessary from a public policy standpoint, though not without error. Subsequent economic theorists sought to eradicate satisfaction measurement and comparisons as value laden and unnecessary (Hicks 1939; Robbins 1938). Recently, satisfaction has again emerged as a basis for making meaningful comparisons across people and products. Virtually all policy recommendations require comparisons of welfare which is proof enough that they are possible (Scitovsky 1951). The important question has become how comparisons of satisfaction or well-being are and should be made (Hammond 1991; Jorgenson 1990; Sen 1979; Simon 1974; Tinbergen 1991).

Our interest is specifically with customer satisfaction. Economic theorists are more often concerned with comparisons of more global economic well-being, which includes not only customer satisfaction but job satisfaction and income evaluation (Poiesz and Grumbkow 1988). Broad based comparisons of customer satisfaction are not exactly new. Andreasen and Best (1977) report meaningful comparisons of customer satisfaction and complaint behavior across a variety of product and service categories, while Pfaff (Lingoes and Pfaff 1972; Pfaff 1977) has used subjective measures of satisfaction to construct an index for the purpose of comparing various food product categories. Wikström (1983) has even compared subjective levels of customer satisfaction across countries (Sweden and the U.S.) and argued that the observed differences can be traced to underlying differences in market performance between the two countries.

Yet some policy researchers have concluded that subjective measures of customer satisfaction are incapable of revealing any meaningful differences (Hunt 1988; Ölander 1988). Most notably, Ölander (1977a, 1977b, 1988) argues persuasively that subjective measures of customer satisfaction are fraught with problems. Foremost among these is the so-called "happy slave" problem. Because customers adapt to the levels of product and service performance available to them, no meaningful differences in satisfaction should emerge. Individual differences in the degree of adaptation within and across industries further compounds the problem. Other problems include the notion that customers may have different yardsticks by which they judge satisfaction. Even if they used the same standards, consumers may have very different product or service alternatives available to them, and/or differ in their knowledge of these alternatives. Finally, customers may fail to express true dissatisfaction or strategically express false dissatisfaction in hopes of receiving some retribution.

At some level Ölander's concerns are very real, as when one might compare satisfaction between customers in relatively wealthy and impoverished countries. At the same time, several considerations suggest that the concerns expressed over comparing

subjective measures of customer satisfaction are overstated. As Pfaff (1977) argues, who is in the best position to evaluate customer satisfaction but customers themselves? People are more similar than different, and these similarities are growing in an increasingly "global" economy. We live in an information age in which consumers from different countries and socio-economic strata of our society are increasingly aware of alternative products and services available in the marketplace.

Equally if not more important are recent advances in satisfaction survey methods and modeling which facilitate our ability to compare subjective measures of satisfaction. Sweden's Customer Satisfaction Barometer (SCSB) embodies these advances and is serving as the prototype for the American Customer Satisfaction Index (ACSI).

THE SWEDISH CUSTOMER SATISFACTION BAROMETER

Established in 1989, the SCSB was the first national customer satisfaction index for domestically purchased and consumed products and services (Fornell 1992). The index is constructed using survey measures obtained from representative customers in each of 32 major Swedish industries which themselves represent approximately 70% of Sweden's gross domestic product (GDP). Those companies that account for approximately 70% of combined industry sales are selected to represent each industry. In cases where a company sells multiple products or services, the "flagship" brand (the product or service with the highest sales in kronor) is chosen to represent the company. For example, Saab Scania is represented by the sales of its 9000 series automobiles while banks are represented by their money lending activities. Each year approximately 100,000 customers are contacted by telephone and screened to obtain a sample that has experience with the products and services in the index. The number of customers who pass the experience screen and agree to participate is approximately 25,000 each year. Industry level sample sizes range from about 250 to over 4000 depending on the number of competitors.

Comparability in the Model

The first step in assuring comparability in the SCSB involves the choice of satisfaction related constructs and how they are modeled. Survey respondents are asked a variety of questions to operationalize five key constructs: (1) customer perceptions of product or service performance, (2) their expectations regarding performance, (3) customer satisfaction, (4) whether they have complained ("voice"), and (5) customer loyalty. The SCSB model, which is presented in Figure 1, posits six relationships among these variables. These relationships are relatively universal in that they cut across all of the products and services in the barometer and are described briefly here (for more extensive descriptions and discussion see Fornell 1992).

- insert Figure 1 here -

Satisfaction is posited to be a function of two antecedent variables, perceived performance or quality and customer's expectations regarding performance (Fornell and Johnson 1993; Johnson and Fornell 1991). Customer satisfaction should increase with the degree to which a product or service provides net benefits that customers value (i.e., perceived performance). Because expectations embody past quality or performance information, they too should positively affect satisfaction. Expectations serve to anchor overall evaluations of satisfaction in the vicinity of the expectations (Oliver 1977, 1980). The size of this anchoring effect depends on the relative strength of the expectations versus performance information (Johnson, Nader, and Fornell 1994). As experienced customers can predict, to some degree, what levels of performance they will receive, expectations should also show a positive relationship to perceived performance.

There are two primary behavioral consequences of satisfaction. Increased customer satisfaction should reduce the incidence of customer voice or complaining behavior. Satisfied customers are also loyal customers, which is the key to the satisfaction-profitability linkage (Anderson, Fornell and Lehmann 1994). Finally, voice may increase loyalty. The size of this relationship reflects the degree to which customers

are allowed to voice their complaints and a firm's ability to address these complaints. That is, the relationship is positive when a firm can turn a complaining customer into a loyal customer. Overall the relationships in Figure 1 are well supported and appear to generalize across Swedish industries (Fornell 1992).

Comparability in Satisfaction Survey Items

The next step in assuring the comparability of satisfaction in the SCSB is to use a survey instrument whose questions are themselves universally applicable and help control for industry differences. This is quite different from what typically occurs in the context of a particular product category or industry where perceived performance is operationalized using customer ratings of a product or service on quality dimensions or attributes that are idiosyncratic to the industry (e.g., attributes of an automobile). In the SCSB, performance is operationalized using two measures of perceived value, the customer's perception of quality received relative to the price or prices paid (benefits relative to costs) and their perception of the price or prices paid relative to quality received (costs relative to benefits). Research demonstrates that this "value" is a common denominator that consumers use to compare even very dissimilar or "noncomparable" products and services (Johnson 1984, 1989). Using value perceptions to measure performance also controls for differences in income and budget constraints across respondents (Hauser and Shugan 1983; Lancaster 1971) which allows us to compare very high and very low priced products and services.

Satisfaction is also surveyed using comparable items. These include the customer's rating of overall satisfaction, how well the product performs relative to an ideal product or service in the industry, and whether performance fall short of or exceeds customer expectations. Theoretically, all three of these rating should reflect the underlying level of satisfaction independent of the particular product, firm, or industry involved (Johnson 1994; Johnson, Anderson and Fornell 1994). Customer voice is, meanwhile, measured in two ways: the incidence of formal complaints to company or

agency managers, and the incidence of informal complaints to personnel or service providers. As for performance and satisfaction, both measures are flexible and apply to a variety of organizations.

Finally, customer loyalty is measured using questions regarding repurchase likelihood and sensitivity to price increases. While very applicable to competitive product and service industries, these loyalty measures are more problematic in the case of government agencies and monopolies. The solution used in the SCSB is to make the questions hypothetical. That is, assuming some other organization could provide an agency's services, how likely would you be to use the agency again and how much more would the agency have to "charge you" before you would switch to the hypothetical competitor? Over time, these questions are becoming less hypothetical and more realistic as government agencies are being reinvented and subjected to increased competition and market pressures.

Comparability in Satisfaction Model Estimation

The third step in assuring the comparability of subjective satisfaction centers on just how the survey items described above are used to operationalize the constructs and estimate the relationships in Figure 1. An important aspect of the Swedish index is that satisfaction (as well as performance, voice, and loyalty) is operationalized as a latent variable within a system of equations. Johnson and Fornell (1991) argue that satisfaction, as a theoretical concept, is a common denominator on which very different people and products may be compared. As a latent theoretical construct, satisfaction is empirically measurable as a weighted average of multiple satisfaction indicators. As Ölander and others have argued, any individual rating or measure which uses a particular yardstick is at best an indirect proxy for satisfaction. Operationalizing satisfaction as the shared variance among a set of multiple satisfaction survey measures provides a more direct measure of latent satisfaction.

This latent variable is estimated with a system of equations, or causal model framework, using the SCSB model in Figure 1. The particular estimation method used to operationalize latent satisfaction and estimate the model is partial least squares or PLS (Fornell 1989; Lohmöller 1989; Wold 1982). PLS is an iterative estimation procedure that corrects for routine least-squares measurement problems and does not impose distributional assumptions on the data. This is particularly attractive in a satisfaction context where distributions are often highly skewed. As a result, PLS is better suited to causal model estimations involving small samples than is, for example, covariance structure analysis using LISREL. It also allows the researcher to operationalize latent variable scores and hence calculate an index value.

Another important feature of PLS is that it aims to explain variances at the observed (measurement) level while LISREL aims to account for observed covariances. In Figure 1, latent satisfaction should ultimately explain variance in loyalty across customers. PLS weighs the individual satisfaction survey items in the satisfaction index so as to maximize the index's ability to explain loyalty. This, in turn, provides a satisfaction index that is comparable in the following sense. In each industry, the satisfaction index explains an endogenous, dependent variable that is universally applicable across industries. The satisfaction index is itself explained by two antecedents that should affect satisfaction in a similar fashion across industries. If the satisfaction index behaves as it should behave according to the model in Figure 1, then its validity and value as a benchmarking source is supported.

Empirical Evidence of Comparability

The ultimate test of the "happy slave" problem and other questions raised regarding the comparability of satisfaction is an empirical one. A recent study by the author and Claes Fornell (Fornell and Johnson 1993) using the SCSB data explicitly examines this issue. In the study we argue that if one can explain differences in satisfaction across industries using some underlying difference in the industries, then the

observed differences are meaningful. Specifically, we argue that product or service differentiation in an industry is one logical basis for explaining differences in expectations, performance, and satisfaction across industries.

Differentiation, in this context, refers to the availability of predictably different options to customer. Differentiated industries offer predictably different options that more directly meet the needs of a heterogeneous population of customers. In contrast, undifferentiated industries offer "no choice." The automobile industry in Sweden is, for example, highly differentiated. Customers can choose among a wide variety of options and are confident in their ability to evaluate differences among them. Police, telecommunications, and public postal services are, in contrast, relatively undifferentiated due to the lack of variety from which to choose. At a more intermediate level of differentiation are banks and insurance companies, where alternatives exist yet customers have difficulty judging their differences. The study found that the level of differentiation across the industries explained fifty-percent of the variance in aggregate perceived industry performance. This performance, in turn, explained over half of the variance in aggregate industry customer satisfaction.

This study has important implications for government agencies who must now benchmark their customers' satisfaction to that observed in private sector industries. In the past, public utilities, monopolies, and government agencies had no competitors on which to benchmark satisfaction levels. Because the industry level differences in satisfaction are meaningful, national indices such as the SCSB provide these agencies and firms with useful benchmarks. The Satisfaction Index scores for the Swedish industries are presented in Table 1. Government owned industries include the pharmacies, local police services, business post, public post, railroads, business telecommunications, public telecommunications, and state sponsored TV broadcasting. To illustrate the differences among industry types, the industries in Table 1 were grouped into three classes: (1) products and product retailers, (2) services, and (3) government owned agencies and

businesses. The average satisfaction indices for each of these three groups from 1989 to 1993 are plotted in Figure 2.

- insert Table 1 and Figure 2 here -

The figure illustrates several interesting points. First, following Fornell (1992; Fornell and Johnson 1993), products and product retailers show systematically higher levels of satisfaction than do competitive services and government owned agencies and businesses. Both of the latter groups are service-oriented, which makes it inherently more difficult to meet specific customer needs. While products meet customer needs largely through their physical means of production, the production of services involves more of the human resources of the firm and customers themselves. This creates greater heterogeneity, on average, in the production of services versus products and lower average performance (Fornell and Johnson 1993; Zeithaml et al. 1988). In Figure 2, products and product retailers show the highest satisfaction and it stays relatively stable over time. Competitive services are below the products and retailers, which is consistent with the nature of service production. The drops in service satisfaction in 1992 and 1993 are due primarily to the recent poor performance in the banking sectors. Finally, the government owned agencies and businesses are generally lowest in satisfaction. This is due both to their service orientation, which makes it difficult to provide consistent quality, and monopoly positions, which limits customer choice.

More important from a benchmarking standpoint is the steady increase in satisfaction for the government sector over the five years in which the index has been in operation. Average satisfaction has increased from 54 to 61 (on a 0 to 100 scale) in this five years. Some of this increase is due to the addition of the high performing state pharmacies in 1990. Even without the pharmacies, however, there is a steady increase in this sector (from 54 to 59). Importantly, the differences between competitive and government owned services is decreasing over time. Following Fornell and Johnson (1993), this suggests that competitive services provide government agencies with a useful

benchmark for industry satisfaction that were not previously available. It appears that government agencies in Sweden are using these attainable benchmarks to improve performance. It would be more unrealistic to expect agencies, on the whole, to achieve the satisfaction levels that we observe for competitive products where the means of production is quite different. A second implication is that individual government owned or regulated businesses, such as the pharmacies, are capable of achieving even higher satisfaction levels. Overall, the SCSB results thus provide government agencies in Sweden with both attainable benchmarks and role models for setting satisfaction goals.

THE AMERICAN CUSTOMER SATISFACTION INDEX

The SCSB serves as the prototype for the American Customer Satisfaction Index (ACSI) which will be released for the first time in October of 1994. The ACSI is a quarterly, national index of customer satisfaction. Sponsored by the University of Michigan, the National Quality Research Center at the Michigan Business School, and the American Society for Quality Control, the index will, in its first year, survey approximately 50,000 customers of approximately 200 companies and government agencies which comprise about 49 percent of U.S. Gross Domestic Product.

There are important differences between the ACSI and the existing SCSB. First, the ACSI is larger in scope given the greater size of the U.S. economy. American firms are also more diverse in that a single firm is more likely to compete in multiple industry sectors. Sampling is, therefore, being done at the "firm level" rather than the "product or service" level. Finally, the ACSI (and future versions of the SCSB) include an expanded set of survey items. In addition to the original SCSB questions, customers will be asked both their expectations and perceptions of performance regarding two key quality components: (1) "fitness for use," or the degree to which a product or service provides those things that the customer personally requires from the product or service, and (2) "things gone wrong," or the degree to which a product or service is free from defects.

Both factors are germane to quality across all U.S. industries and will provide interesting bases for comparison.

Conclusions: Implications for Government Agencies

Our recent experience in the development of national customer satisfaction indices illustrates a number of important principals and concepts that should help government agencies as they actively implement satisfaction measurement systems. The first is that there is a relatively "universal" model of the antecedents and consequences of customer satisfaction. The challenge that agencies face is in translating the constructs in Figure 1 to the particular agency context. Customer loyalty, for example, may be "repurchase" in some agencies (e.g., the buying of Census Bureau data) and "compliance" in others (e.g., with an IRS regulation or rule). A second implication is that there are universal ways of asking the survey questions needed to operationalize such things as performance, satisfaction, and loyalty. This involves a focus on common denominators, such as "value" when operationalizing perceived performance, and using multiple standards of comparison, as when measuring satisfaction.

Once a flexible model and a set of survey measures are in place, the measures should be used to develop indices of the key constructs. This is especially true for satisfaction where any single survey item is at best a proxy for a customer's overall evaluation of their experience with a firm or agency. Ideally, the satisfaction index should be estimated within the context of a model (e.g., Figure 1) where, for example, performance and expectations explain satisfaction and satisfaction, in turn, explains customer voice and loyalty. These steps address many of the criticisms raised by consumer and policy researchers such as Ölander and Hunt toward the use of subjective measures of satisfaction. They help assure comparability in satisfaction measures across people and industries.

However, the ultimate test of this comparability is an empirical one. As the Swedish experience shows, customer satisfaction is empirically comparable. When

customer satisfaction is properly surveyed, measured and modeled, it allows one to compare "apples and oranges." The resulting comparisons provide useful benchmarks for government agencies as they improve quality. Competitive service industries provide a very straightforward benchmark that, based on the Swedish experience, appears attainable for government agencies as a whole. A national index also allows one to identify a particular agency or agencies to serve as role models and provide even higher satisfaction goals.

At another level, having established the comparability of satisfaction surveys, government agencies can use satisfaction index results to make better decisions and resource allocations. Existing productivity measures and price indices are limited in the way they account for quality changes (National Economic Research Associates 1991). Resources could be allocated more effectively by targeting industries or agencies that rate particularly low on satisfaction to help improve overall consumer welfare. For example, if the IRS rates particularly low on satisfaction, allocating resources toward improving customer satisfaction should more than pay for itself in terms of increased efficiency, compliance and resulting revenue generation. Finally, agencies will benefit by having a more complete picture of their organizations. Any comprehensive strategic plan for a public or private organization must integrate the organization's goals for achieving customer, employee, and owner satisfaction. In government agencies, taxpayers are the ultimate "owners." As agencies strive to meet customer needs, build customer loyalty, and save taxpayer dollars, these owners are the ultimate winners.

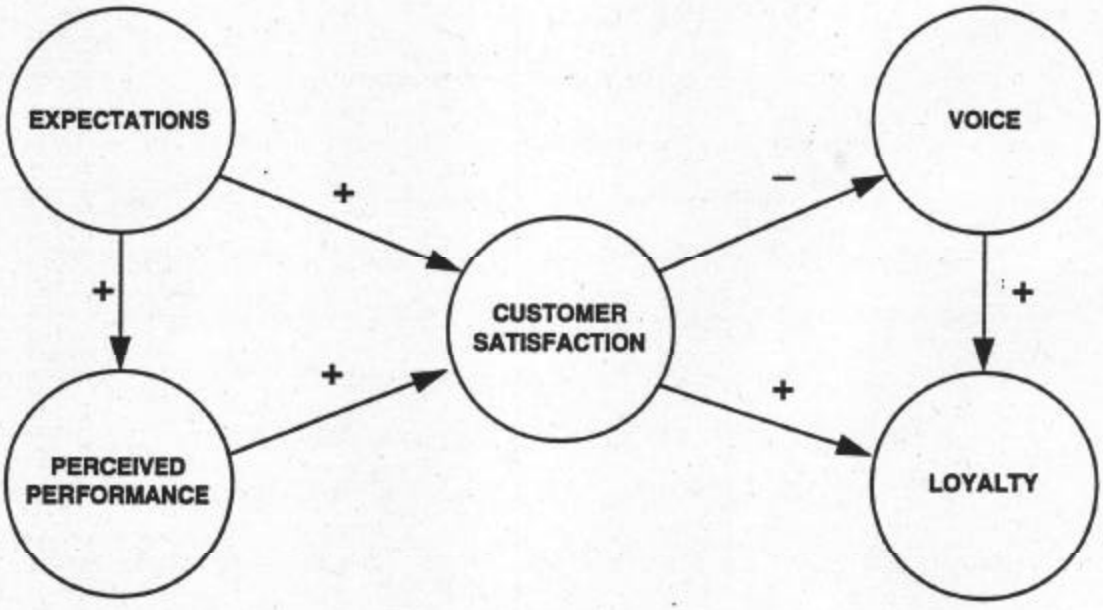


Figure 1. The SCSB Model

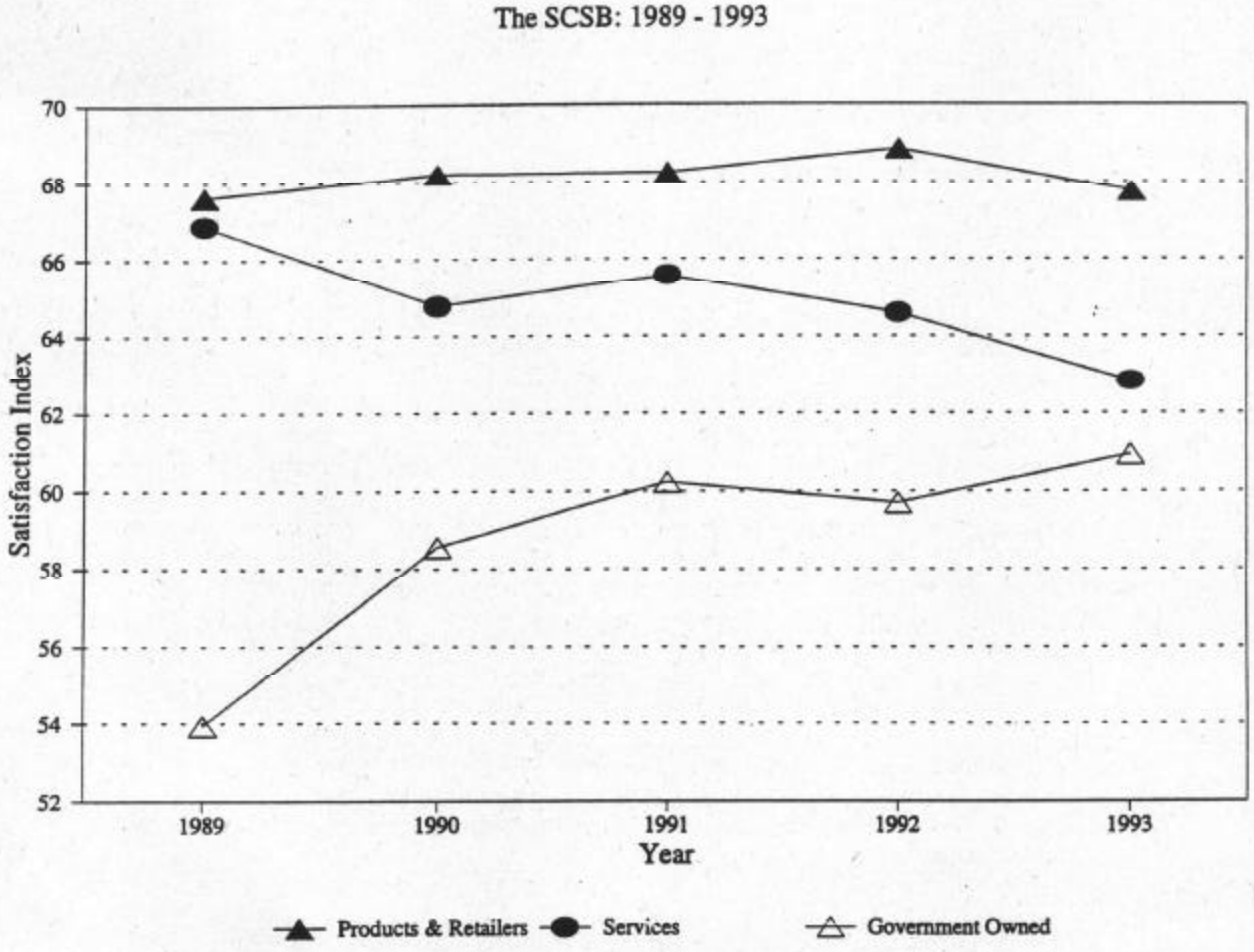


Figure 2. Year-to-Year Changes in the SCSB

Industry	Satisfaction by Year*				
	1989	1990	1991	1992	1993
Airlines	67	67	68	63	65
Automobiles	77	78	78	76	77
Banks (Business)	70	67	64	65	59
Banks (Public)	69	69	67	67	63
Clothing Retailers	63	63	62	63	63
Computers (Main Frames)	69	63	63	64	62
Computers (Business PCs)	70	66	66	67	64
Department Stores	62	63	61	61	60
Food Processors**	68	71	71	72	70
Furniture Retailers	64	63	65	65	64
Gas Stations	67	68	70	70	70
Grocery Stores	66	68	65	67	66
Insurance (Business)	64	62	64	62	61
Insurance (Automobile)	66	63	66	64	62
Insurance (Life)	65	65	63	61	54
Mail Order	na	64	63	64	64
Newspapers	na	60	64	63	62
Pharmacies	na	76	73	72	74
Police	56	55	58	59	58
Postal Service (Business)	59	62	65	61	66
Postal Service (Public)	65	61	67	63	65
Railroad	44	55	54	54	54
Shipping	na	65	69	67	69
Travel (Charter)	68	67	68	68	68
Telecom. (Business)	53	57	57	61	61
Telecom. (Public)	55	59	61	59	61
TV Broadcasters	44	43	47	48	49

* Satisfaction index is on a 0 to 100 scale.
** The averages for Food Processors include six separate food industries (basic foods, candy and coffee, baked goods and dairy products, beer, meat products, and canned and frozen foods).

Table 1. Swedish Customer Satisfaction Barometer Results

REFERENCES

Anderson, Eugene W., Claes Fornell, and Donald R. Lehmann (1994), "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden," *Journal of Marketing*, forthcoming.

Andreasen, Alan R. and Arthur Best (1977), "Consumers Complain - Does Business Respond?," *Harvard Business Review*, July-August 1977, 93-101.

Bentham, J. (1802), in E. Dumont, *Theory and Legislation*, London: Trubner, 1871.

Fornell, Claes (1989), "The Blending of Theoretical and Empirical Knowledge in Structural Equations with Unobservables," in H. Wold (ed.), *Theoretical Empiricism*, New York: Paragon House, 153-173.

_____ (1992), "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing*, 56 (January), 6-21.

_____ and Michael D. Johnson (1993), "Differentiation as a Basis for Explaining Customer Satisfaction Across Industries," *Journal of Economic Psychology*, 14 (4), 681-696.

Hammond, Peter J. (1991), "Interpersonal Comparisons of Utility: Why and How they are and Should Be Made," in J. Elster and J. E. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, Cambridge: Cambridge University Press, 200-254.

Hauser, John R. and Steven M. Shugan (1983), "Defensive Marketing Strategies," Marketing Science, 2 (4), 319-360.

Hicks, J. R. (1939), *Value and Capital*, London: Oxford University Press.

Hunt, H. Keith (1988), "Consumer Satisfaction/Dissatisfaction and the Consumer Interest," in E. Scott Maynes (ed.), *The Frontier of Research in the Consumer Interest*, Columbia, MO: American Council on Consumer Interests, 731-747.

Johnson, Michael D. (1984), "Consumer Choice Strategies for Comparing Noncomparable Alternatives," *Journal of Consumer Research*, 11 (December), 741-753.

_____ (1989), "The Differential Processing of Product Category and Noncomparable Choice Alternatives," *Journal of Consumer Research*, 16 (December), 300-309.

_____ (1994), "The Four Faces of Aggregation in Customer Satisfaction Research," Working Paper, Ann Arbor, MI: National Quality Research Center.

_____, Eugene W. Anderson and Claes Fornell (1994), "Rational and Adaptive Performance Expectations in a Customer Satisfaction Framework," Working Paper, Ann Arbor, MI: National Quality Research Center.

_____ and Claes Fornell (1991), "A Framework for Comparing Customer Satisfaction Across Individuals and Product Categories, *Journal of Economic Psychology*, 12 (2), 267-286.

_____, Georg Nader and Claes Fornell (1994), "Expectation, Perceived Performance, and Customer Satisfaction for a Complex Service: The Case of Bank Loans," Working Paper, Ann Arbor, MI: National Quality Research Center.

Jorgenson, D. W. (1990), "Aggregate Consumer Behavior and the Measurement of Social Welfare," *Econometrica*, 58, 1007-1040.

Lancaster, Kelvin (1971), *Consumer Demand: A New Approach*, New York: Columbia University Press.

Lingoes, James C. and Martin Pfaff (1972), "The Index of Consumer Satisfaction: Methodology," in M. Venkatesan (ed.), *Association for Consumer Research: 3rd Annual Conference Proceedings*, 689-712.

Lohmöller, J.-B. (1989), *Latent Variable Path Modeling with Partial Least Squares*, Heidelberg: Physica-Verlag.

Meeks, J. G. Tulip (1984), "Utility in Economics: A Survey of the Literature," in Charles F. Turner and Elizabeth Martin (eds.), *Surveying Subjective Phenomena, Volume 2*, New York: Russell Sage Foundation, 41-91.

National Economic Research Associates (1991), *Developing a National Quality Index: A Preliminary Study of Feasibility*, Report prepared for The American Quality Foundation, White Plains, NY: NERA, Inc.

Ölander, Folke (1977a), "Can Consumer Dissatisfaction and Complaints Guide Public Consumer Policy?" *Journal of Consumer Policy*, 1 (Spring), 124-137.

_____ (1977b), "Consumer Satisfaction - A Skeptic's View," in H. Keith Hunt (ed.), *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, Cambridge, MA: Marketing Science Institute, 409-452.

_____ (1988), "Consumer Satisfaction/Dissatisfaction and the Consumer Interest," in E. Scott Maynes (ed.), *The Frontier of Research in the Consumer Interest*, Columbia, MO: American Council on Consumer Interests, 753-759.

Oliver, Richard L. (1977), "Effect of Expectation and Disconfirmation on Post-exposure Product Evaluations: An Alternative Interpretation," *Journal of Applied Psychology*, 62, 480-486.

_____ (1980), "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, 17 (November), 460-469.

Pfaff, Martin (1977), "The Index of Consumer Satisfaction: Measurement Problems and Opportunities," in H. Keith Hunt (ed.), *Conceptualization and Measurement of Consumer Satisfaction and Dissatisfaction*, Cambridge, MA: Marketing Science Institute, 36-71.

Poiesz, Theo B. C. and Jasper von Grumbkow (1988), "Economic Well-Being, Job Satisfaction, Income Evaluation and Consumer Satisfaction: An Integrative Attempt," in W. F. van Raaij, G. M. van Veldhoven and K. E. Wärneryd (eds.), *Handbook of Economic Psychology*, Dordrecht, the Netherlands: Kluwer Academic Publishers, 570-593.

Robbins, L. (1938), "Interpersonal Comparisons of Utility: A Comment," *Economic Journal*, 48, 635-641.

Scitovsky, Tibor (1951), "The State of Welfare Economics," *American Economic Review*, 41, 303-315.

Sen, A. K. (1979), "Interpersonal Comparisons of Welfare," in M. J. Boskin (ed.), *Economics and Human Welfare: Essays in Honor of Tibor Scitovsky*, New York: Academic Press, 183-201.

Simon, John L. (1974), "Interpersonal Welfare Comparisons can be Made - and Used for Redistribution Decisions," *Kyklos*, 27, 63-98.

Tinbergen, Jan (1991), "On the Measurement of Welfare," *Journal of Econometrics*, 50 (1), 7-13.

Van Raaij, W. Fred (1981), "Economic Psychology," *Journal of Economic Psychology*, 1 (1), 1-24.

Wikström, Solveig (1983), "Another Look at Consumer Dissatisfaction as a Measure of Market Performance," *Journal of Consumer Policy*, 6 (1), 19-35.

Wold, Herman (1982), "Systems Under Indirect Observation Using PLS," in Claes Fornell (ed.), *A Second Generation of Multivariate Analysis: Methods*, New York: Praeger, 325-347.

Zeithaml, Valerie A., Leonard L. Berry, and A. Parasuraman (1988), "Communication and Control Processes in the Delivery of Service Quality," *Journal of Marketing*, 52 (April), 35-48.

Session 3 Customer Surveys

Discussion

Robert M. Groves
University of Michigan and Joint Program in Survey Methodology

In addressing the notion of customer service standards and customer measurement programs, the U.S. government is attempting to import a set of ideas tried in the commercial sector. It is useful to note that the ideas, once tried, do not always prove themselves to be cures of the ailments of modern commercial organizations. Why they work sometimes and why they don't other times is the topic of much current debate. We are now living through the period of time when most of you in the audience are determining whether this is the management philosophy of the week or the beginning of a new perspective on agency functioning.

First, let's take a minute to review the recent history of the commercial sector. the common lessons of the "customer satisfaction movement" are:

- external threats help shock organizations into paying attention to their customers
- customer orientation succeeds only when top management forces it, repeatedly, in every forum, relentlessly
- measurement of satisfaction only once is nearly useless
- measurement of satisfaction without simultaneous measurement of production/service activities related to satisfaction is nearly useless

Now let's see whether these lessons are relevant to the papers we have heard presented today.

The Devens Paper

There are really three parts to the Devens paper -- a commentary on the customer-orientation movement, the description of a survey, and a review of the feedback loop to managers.

Commentary on customer-orientation movement

Devens notes that managers do "want" to improve their operations, one way or another. Clearly, the question is whether the operations *are* being improved. It seems clear that all change is difficult to induce in government agencies, but if anything, continuous improvement changes may be more difficult in ongoing statistical operations than in other areas. The problem stems from the need to maintain comparable measurement

systems over time in ongoing series. The concern is change that affects the bias properties of estimates, not just the variance properties (yet even changes that theoretically affect only variance properties (eg. a new sample) can affect bias).

Why is that a concern? Rarely do we have information that bias change goes in the right direction. Exceptions are the higher victimization rates of the NCS and higher unemployment rates of CPS, where there is a strong model of tendencies to underreport those phenomena, but even there the model can be easily challenged.

So incremental change in statistical operations may be harder than incremental change in other fields because the product of today has more value if it is comparable to the product of yesterday.

Devens notes that many managers are skeptical about customer focus because customers aren't qualified to judge the quality aspects of statistical series. This comment on the surface sounds familiar to those in charge of the design of the 1994 Chevrolet Caprice, but statistical agencies may have greater challenges than manufacturers of other products. The General Motors managers did have access to many marketing research studies about the concerns and interests of their customers, but apparently discounted them. Most statistical agencies have no equivalent of the market research function, and thus customer desires are only indirectly and erratically communicated. As with automobiles it is easy to confuse the fact that only the customers know what information they need, but only the statisticians may know how best to produce it.

Description of the Survey

In this section there is very little concern about issues concerning the sampling frame and inference. This unfortunately is a serious lacuna in most of the literature about the Total Design Method. Despite its name it does not address issues of coverage error in surveys. These issues are complex and largely uncharted in surveys of customers of statistical agencies because the target population of customers has usually not be fully enumerated at the time of a survey. Even the definition of "customer" becomes a complex one, when considering information as a product.

The most important point of this section is that careful planning of a mailed questionnaire can yield high response rates. For this reason alone, this is an important paper. When government agencies are telling others and are being told that response rates in the 20% range are the highest to be expected, this work has shown that careful planning and execution can obtain high participation rates. High response rates are doubly important in this area, because of the finding that nonrespondents to satisfaction surveys tend to be dissatisfied with the service or product.

Review of Feedback Loop to Managers

Despite some commentary that the survey did not yield clear findings, conclusions were indeed drawn from the data collection -- the need for more timely faster products, one stop service for questions, and clearer presentations of information. Of these, it appears that the organization has addressed timeliness of products most directly. It is noteworthy

that this area was not the one with the largest gap between expectations and perceived performance. This might be an example of management either ignoring the empirical findings or management supplementing the empirical work with other external information about performance. The paper does not reveal which is the appropriate interpretation.

The Johnson Paper

This paper describes a large effort to construct a useful measure of consumer satisfaction across all sectors of the economy. It is conceptualized as another macroeconomic indicator, measuring an outcome of production-- in one term "post-consumptive utility" The paper is divided into three sections: a) Can satisfaction be compared across sectors/industries?; b) a description of the Swedish Customer Satisfaction Barometer; and c) an announcement of the American Customer Satisfaction Index

Can satisfaction be compared across sectors/industries?

How would you know whether you'd have the answer to this question? Would it depend on the ability to predict behavior? What behavior? The evidence of comparability presented includes the finding that 50% of the variance in perceptions of performance is explained by the amount of differentiation in the industry, and performance explains 50% of variance in satisfaction. Clearly, one would like to assemble more evidence: behavioral outcomes measured on same persons over time, stable relationships between satisfaction levels and growth rates, complaint rates, etc.;

Swedish Customer Satisfaction Barometer

This is a large data collection and estimation series, running since 1989, covering 70% of sales in each industry, and measuring one product or service per company. Clearly the process of sampling firms and products/services is a nontrivial problem, as those in the consumer and producer price index measurement process know. The sampling problem facing this index is multi-level (sectors, firms, products/services, customers) there are important sources of variance at each level and important sources of information about customer satisfaction.

The concepts measured include: perception of performance, expectations regarding performance, satisfaction, reports of whether the customer has complained about the product or service, and customer loyalty. These are difficult measurement issues. For example, the approach is forced to use hypothetical questions on loyalty, using words like "if another agency could provide the same service."

The paper presents findings from the Swedish effort that are stimulating, given the current effort at measuring customer satisfaction in U.S. government agencies. For example, there is the finding that Swedish government agencies (police, pharmacies, post office, railroads, telecommunications, tv broadcasting) started with lower satisfaction and rose in satisfaction faster than other sectors. One wonders whether that finding will be duplicated in the US. The finding itself illustrates one of the challenges to the measurement process. To what extent is measurement of satisfaction with services of a government agency affected by general feelings of civic pride, trust in government,

political efficacy?

American Customer Satisfaction Index

The last section of the paper sketches out the plans for a U.S. customer satisfaction index. This effort will be different from the Swedish experience in that the U.S. population of firms offer more diverse products and services.

In both this and the Swedish index description there seems to be most emphasis placed on the psychometric properties of the measurement and little concern with traditional survey issues of coverage of the target population (telephone surveys are planned), nonresponse error, and measurement errors associated with social desirability, mode of data collection, etc.

Summary

These two papers, although seemingly disparate in the focus, can serve to remind us of two important debates in customer satisfaction:

- Is satisfaction merely a function of the difference between reported expectations and performance ratings?

- Do expectations cause perceived performance?

Let me summarize my reactions to the papers:

1. In government agencies, we are at the beginning of the customer measurement process. Its value rests on repeated measurement, empirical assessment of relationship between actions of employees and satisfaction, and change in satisfaction over time. We are a long way from this status of measurement and innovation.

2. Both of these papers appear to miss the connection to actual activities of the units to increase satisfaction. They are more heavily focused on measurement than how measurement can lead to change and then later to improved satisfaction.

3. The papers flow from different conceptual bases; a debate that is not joined by the two. One stems from the notion that satisfaction is in some sense the gap between expectations and performance. The other attempts to add another concept, perceptions of an ideal service or product, in order to calibrate the gap between expectations and performance. These conceptual differences are part of the debate now ongoing in the satisfaction measurement field. These are important issues for the practical import of satisfaction measurement. If, for example, performance at time 1 sets expectations at time 2, then poor performance lowers expectations, and in one perspective, would yield higher satisfaction, as expectations and performance were in sync. From the other perspective, departures between performance and the concept of the ideal, would be larger at time 2 and lead to large "performance gaps."

If government agencies take seriously the measurement of customer satisfaction, they will

inevitably be forced to attend to such issues. They are key to the meaningful tracking of how satisfaction can change with improved performance of organizations. We are at the beginning of this process for government agencies, and we are in the debt of these two papers for alerting us to such issues.

DISCUSSION

Elizabeth Martin
Bureau of the Census

Both papers presented in this session raise issues which are relevant to the current efforts by federal agencies to respond to a presidential order to survey their customers, measure satisfaction, and use the results to set service standards and provide customers with greater choice in services. Significantly, the executive order further states that "as information about customer satisfaction becomes available, each agency shall use that information in judging the performance of agency management and in making resource allocations."

If this aspect of the executive order comes to pass, then the issue of the comparability of customer satisfaction measurements among agencies and across diverse products and services is of more than academic interest. I'd like to start by focussing on the issue of comparability of measurement as addressed in the Johnson paper, which describes a customer survey conducted across 32 industries in Sweden, including government-run industries such as the postal service and railroads. I'll be drawing on my recent involvement in an effort to design a generic customer satisfaction questionnaire for use by all of the agencies of the Department of Commerce. Next, I'll discuss the Devens paper, which discusses a customer satisfaction survey targetted much more narrowly, to users of employment and unemployment statistics produced by the Bureau of

Labor Statistics. It raises some interesting issues about the utility and consequences of customer satisfaction surveys.

Johnson and his colleagues are concerned with a very ambitious effort to develop a customer satisfaction index (the Swedish Customer Satisfaction Barometer) which can be applied across industries, products, and, ultimately, in different countries. Customers were sampled from 32 Swedish industries, including several which were government-run, which represented 70 percent of Sweden's gross domestic product. The companies that accounted for approximately 70 percent of combined industry sales were selected to represent each industry. Each company was represented by its product or service with the highest sales. Each year, 100,000 customers were contacted by telephone and screening questions were asked to determine if they had experience with the products and services chosen to represent the sample companies. About one quarter were eligible and were asked questions to measure their perceptions of performance, their expectations about performance, their satisfaction, whether they ever complained, and their loyalty.

Johnson argues that the survey measurements may be used to compare satisfaction and performance across industries on the following basis: that the model of customer satisfaction which informs their measurements is universally applicable across industry; the measurements used in their survey are universally applicable across all industries; and that there exist meaningful differences in satisfaction between industries which can be explained by industry-level differences in degree of product or

service differentiation--the a[illegible]t of choice offered customers. He also suggests that that the index provides meaningful and comparable information about customer satisfaction for government agencies and industries as well as private sector ones.

Although it may be possible to design a survey which yields comparable measurements of satisfaction for customers of diverse products and services produced by different industries, Johnson and his colleagues have not satisfactorily made the case for the comparability of their measurements. They need to address the following questions: First, is the definition of a customer comparable across different industries? Second, are the sampling frame and response rates comparable across industries and over time? Finally, are their satisfaction measurements comparable and meaningful across different industries and over time?

The first issue, of what is a customer, usually is not terribly ambiguous in the private sector, but it bedevils attempts to measure customer satisfaction in a government setting. To most of us, a customer is someone who purchases a commodity or service, usually by choice or voluntarily. In a government setting, many products and services are not purchased directly by their users, but subsidized in whole or in part by taxes. Many government products and services are not received voluntarily on the part of the "user" or recipient. Many "customers" of police services or tax collection services no doubt would, if they had the choice, choose not to obtain the service at all. The Johnson paper does not address this issue. In their survey, they defined as customers persons who had experience with the products and services surveyed.

This is a reasonable strategy, but it is important to include measures in the survey to permit the analyst to separately identify voluntary customers of government services, who obtained a product or service by choice, and involuntary customers, who did not choose to obtain the product or service. Comparisons of the former group with customers of private industry may be meaningful, but comparisons involving the latter group probably would not be.

Another definitional problem arises because many services offered by the government are not intended to benefit those who experience them directly, but to protect or benefit others, such as the public, who may not even be aware of their existence. For example, one service provided by a Department of Commerce agency is the inspection of fisheries. Presumably this service is ultimately intended to benefit fish-eaters by ensuring the quality of fish, and only indirectly benefits the fisheries themselves. In this example, it would be difficult to measure the satisfaction of customers who may be unaware that a service exists, much less that they are recipients of it. This sort of issue can make it difficult to identify who should be regarded as the customers of government services, and this category of "customer" would be ruled out by Johnson's screening criterion of "having experience" with a product or service.

A second set of very difficult issues affecting comparability of data has to do with the identification and sampling of customers. In order to make comparisons across industries, one must be certain that the samples are comparable. Johnson reports that each year customers were identified in surveys of the public in

which 100,000 individuals were asked about their experiences with the target products. It is not clear how or whether the sample represents organizational customers. For many of the industries being evaluated--such as banks, railroads, main frame computers--much if not most of their business would be with other organizations or businesses, not with individual consumers. Organizations, and their experiences as customers, do not appear to be represented in the index. Their absence reduces the meaningfulness of customer satisfaction measures for industries in which organizational customers represent a large share of all customers, and reduces comparability of measures across industries which differ in their customer base.

Even if one accepts the limitation that only individual customers are represented in the sample, it is still unclear what universe the results represent. Johnson surveyed customers of the leading products of companies representing 70 percent shares of each of a set of industries which together accounted for 70 percent of the Swedish GDP. One would expect the companies, products, and industries included in the index to change over time with changes in the economy. This implies that there are two potential sources of change in the value of the index: changes in customer satisfaction for a given set of products, and changes in the composition of products, companies, and industries which make up the index. Given the uncertain interpretation which could be put on any given change in its level, it is not clear how a customer satisfaction index defined this way can provide useful information about trends. Moreover, Johnson's definition appears to leave out most customers,

since 75 percent of his sample was ineligible for the survey. The limited and rather peculiar constraints on the set of customers included in the survey seem to reduce its usefulness as a general index of customer satisfaction with wide applicability across different countries.

The quality and comparability of the results of customer satisfaction surveys depends not only on the quality of the sampling frame, but also on response rates. Johnson presents no information about response rates in the customer survey he reports. In order to make comparisons among industries, one would want reasonably high response rates for all the industries being compared. If response rates varied among industries, then artifactual differences in satisfaction may result from greater nonresponse bias for some than others.

In general, the construction of sampling frames for customer surveys is problematic. If there exist records of purchases, orders, or logs of telephone or other contacts, then these may be used as a sampling frame. However, for many services and products, there are no records which identify customers, especially if no formal or recorded transaction takes place. Customers who pay cash for a product or service, or who listen to the weather station or look up information in a census publication in the library cannot be readily identified. Thus, sampling from records or logs of customer transactions is likely to provide uneven coverage of customers depending on the nature of the industry, how it conducts business with its customers, and the quality and completeness of the records it keeps about customers or transactions. In some

customer surveys, samples are drawn from lists of customers provided by an agency or firm specifically for the survey. Such lists can be very vulnerable to selection bias, since organizational representatives who know that customer satisfaction is to be evaluated are likely to overrepresent satisfied customers in their lists. This selection bias may vary among agencies or organizations and could have a very adverse effect on the comparability of satisfaction measures across agencies or organizations.

The third issue which needs more attention in the Johnson paper is the comparability and meaningfulness of the measurements, especially when applied in a government context. Performance was measured as value, or benefits relative to costs, which seems not to apply very well to products or services which have no specific or direct cost attached to them, as is the case for many government products and services. The key construct of customer loyalty was operationally measured by intention to repurchase and insensitivity to price. Because these measures do not fit government's transactions with its customers, Johnson and his colleagues changed the measures to hypothetical ones for government agencies. It is questionable whether this modification yields results comparable to the original measure. Finally, customer expectations were measured retrospectively, that is, customers were asked to report what their expectations had been at the time of purchase. Retrospective reports of past attitudes are notoriously biased toward present attitudes, and it is highly likely that this measure of "expectations" is contaminated by respondents' subsequent experiences with

a product. This flaw would make it impossible to test the effect of prior expectations on satisfaction or perceived performance.

In summary, customer surveys which aim to compare across diverse industries and products (such as the SCSB discussed by Johnson) potentially are affected by very serious problems of data comparability, including lack of comparability arising from sample design, differential nonresponse, and the measurements themselves. There appears to be a considerable amount of careful methodological and statistical work that still needs to be done to ensure that customer surveys are designed to yield meaningful comparisons of customer satisfaction across industries and over time. Until that groundwork is done, such comparisons should be made cautiously.

The Devens paper raises a different set of issues. The Customer Satisfaction Survey he reports on was much narrower in scope and purpose than the satisfaction index discussed by Johnson, and the issues of data comparability are not nearly as serious, if they exist at all. The survey of customers of employment and unemployment statistics was well done, and obtained an admirably high response rate (88 percent) using reminders and mailings of follow up questionnaires. The survey assessed several aspects of the statistical product, including data quality, ability of staff to answer technical questions, etc. Devens reports that the survey results moved the agency to take several actions to improve the timeliness of the release of its statistics, which the survey showed customers thought was very important.

The paper includes a couple of telling comments by the author, who personally evaluates the survey as "useful but not nearly in

proportion to the skills exercised or the resources expended." He also notes the very low response rate--8 percent-- obtained from a survey of front-line employees upon release of the customer survey, and voices his suspicion that this low rate is due to "failure to convince the front line that the customer satisfaction survey was serious."

Devens' remarks remind us of a couple of key points about customer surveys of this type. First of all, there needs to be clear specification of the goals of the survey, and an understanding of how the information from the survey will be used, in order for the survey to be useful. (This point applies to any survey, not just customer surveys.) To be taken seriously by employees, a customer survey should be designed to address questions to which managers and employees need or want the answers. In the case of the survey Devens reports, it appears that the survey was not credible to managers, and was used in a very limited way by them, reducing the meaningfulness of the survey.

The second point is that customer surveys can themselves affect the expectations of customers and employees. Carrying out a customer survey may raise the expectations of customers (and employees) that a company or agency is going to do something to improve service. If that doesn't happen, and if the survey turns out to be an empty exercise, then the indirect effect of a customer survey may be to reinforce the cynicism of customers or employees or both.

Taking these two points together, and returning to the earlier discussion, we can draw several general conclusions about customer

surveys: Unless an agency or company plans to actually use the results of a customer survey, it shouldn't conduct the survey. The survey should be planned with clear goals and uses in mind. It should be designed to provide fairly specific information that represents useful feedback to managers and employees, and that has implications for action. If the intent is to compare customer satisfaction over time, among products, or among industries and agencies, then the survey should be designed and data evaluated to ensure that results are comparable and can support the comparisons to be made.